

# **APPENDIX A**

## Artificial Protein Cavities as Specific Ligand-binding Templates: Characterization of an Engineered Heterocyclic Cation-binding Site that Preserves the Evolved Specificity of the Parent Protein

Rabi A. Musah<sup>1</sup>, Gerard M. Jensen<sup>1</sup>, Steven W. Bunte<sup>2</sup>,  
Robin J. Rosenfeld<sup>1</sup> and David B. Goodin<sup>1\*</sup>

<sup>1</sup>Department of Molecular Biology, MB8, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla CA 92037, USA

<sup>2</sup>U.S. Army Research Laboratory, Attn.: AMSRL-WM-BD, Aberdeen Proving Ground, MD 21005-5066, USA

Cavity complementation has been observed in many proteins, where an appropriate small molecule binds to a cavity-forming mutant. Here, the binding of compounds to the W191G cavity mutant of cytochrome *c* peroxidase is characterized by X-ray crystallography and binding thermodynamics. Unlike cavities created by removal of hydrophobic side-chains, the W191G cavity does not bind neutral or hydrophobic compounds, but displays a strong specificity for heterocyclic cations, consistent with the role of the protein to stabilize a tryptophan radical at this site. Ligand dissociation constants for the protonated cationic state ranged from 6  $\mu$ M for 2-amino-5-methylthiazole to 1 mM for neutral ligands, and binding was associated with a large enthalpy-entropy compensation. X-ray structures show that each of 18 compounds with binding behavior bind specifically within the artificial cavity and not elsewhere in the protein. The compounds make multiple hydrogen bonds to the cavity walls using a subset of the interactions seen between the protein and solvent in the absence of ligand. For all ligands, every atom that is capable of making a hydrogen bond does so with either protein or solvent. The most often seen interaction is to Asp235, and most compounds bind with a specific orientation that is defined by their ability to interact with this residue. Four of the ligands do not have conventional hydrogen bonding atoms, but were nevertheless observed to orient their most polar CH bond towards Asp235. Two of the larger ligands induce disorder in a surface loop between Pro190 and Asn195 that has been identified as a mobile gate to cavity access. Despite the predominance of hydrogen bonding and electrostatic interactions, the small variation in observed binding free energies were not correlated readily with the strength, type or number of hydrogen bonds or with calculated electrostatic energies alone. Thus, as with naturally occurring binding sites, affinities to W191G are likely to be due to a subtle balance of polar, non-polar, and solvation terms. These studies demonstrate how cavity complementation and judicious choice of site can be used to produce a protein template with an unusual ligand-binding specificity.

Present addresses: R. A. Musah, Department of Chemistry, State University of New York at Albany, 1400 Washington Avenue, Albany, NY 12222, USA; G. M. Jensen, Gilead Sciences, Inc., 650 Cliffside Drive, San Dimas, CA 91773, USA.

Abbreviations used: CCP(MKT), cytochrome *c* peroxidase produced by expression in *Escherichia coli* containing Met-Lys-Thr at the N terminus, Ile at position 53, and Gly at position 152; CCP, cytochrome *c* peroxidase; cyt *c*, cytochrome *c*; W191G, mutant in which Trp-191 is replaced by Gly; MPD, 2-methyl-2,4-pentandiol; WT, wild-type; 2a5mt, 2-amino-5-methylthiazole; 12dmi, 1,2-dimethylimidazole; 2at, 2-aminothiazole; 4ap, 4-aminopyridine; 3ap, 3-aminopyridine; 1mim, 1-methylimidazole; 2ap, 2-aminopyridine; 34dmt, 3,4-dimethylthiazole; 4mim, 4-methylimidazole; im, imidazole; 2mim, 2-methylimidazole; imp, imidazo[1,2-a]pyridine; py, pyridine; bzi, benzimidazole; 345tmt, 3,4,5-trimethylthiazole; 1vim, 1-vinylimidazole; 2a4mi, 2-amino-4-methylthiazole; 3mt, 3-methylthiazole; quin, quinoline; 2eim, 2-ethylimidazole; 234tmt, 2,3,4-trimethylthiazole; tet, tetrazole; pz, pyrazole; 3apz, 3-aminopyrazole; 2mp, 2-mercaptopyrimidine; 3cp, 3-cyanopyridine; iso, isoniazide; anth, anthranilonitrile; 5a1pc, 5-amino-4-pyrazole carbonitrile; 4a5ic, 4-amino-5-imidazole carboxamide; indz, indazole.

E-mail address of the corresponding author: [dbg@scripps.edu](mailto:dbg@scripps.edu)

\*Corresponding author

**Keywords:** protein engineering; protein-ligand binding; cavity complementation; chemical rescue; cavity mutant

## Introduction

Several recent studies have demonstrated the feasibility of creating artificial cavities within protein structures that possess well-defined affinity and specificity for binding small molecules.<sup>1-4</sup> Further characterization of the properties of these artificial binding sites will enhance our understanding of basic features of protein folding and stability. For example, this approach could provide an important advance over mutational analysis of protein-ligand interactions, as a large number of ligands can be used that vary in hydrophobicity, size or shape in a much more subtle way than is possible with amino acid substitutions. Engineered cavities may also have practical uses, such as in the development of specific biosensors, where a designed protein template is used to specifically detect molecules of interest. Finally, a well-defined system consisting of an engineered protein receptor and an array of ligands that bind with known structure and energetics could aid in the development of tools for drug design. Extraordinary efforts are currently underway in the pharmaceutical industry to identify and optimize small molecule binding to specific substrate or effector binding sites.<sup>5</sup> Alternative approaches have been reported in which small molecules are identified that interrupt specific protein-protein interactions.<sup>6,7</sup> For both of these approaches, it is important that binding potential is localized to small regions of the interface. However, it is not entirely clear how universal factors such as hydrophobicity, hydrogen bonding, solvation, electrostatics, and protein conformational dynamics combine to distinguish evolved natural binding sites from more common surface clefts.<sup>8</sup> Even more uncertain is the degree to which cavities created by mutation will resemble natural binding sites with respect to these properties.

In parallel with experimental strategies, computational approaches are providing ever more accurate predictions of protein stability, ligand-binding affinity and conformation. Developments in computational methods such as free-energy perturbation<sup>9,10</sup> and  $\lambda$ -dynamics<sup>11</sup> and in more traditional docking<sup>12</sup> and molecular mechanics simulations<sup>13</sup> have advanced to the point of providing useful tools for screening drug candidates. These methods have been primarily developed and tested using data for ligand binding to naturally evolved sites in proteins and enzymes. While such sites have provided a particularly appropriate benchmark for computational method development, tests of their performance with ligands bound to artificial protein cavities may be of sig-

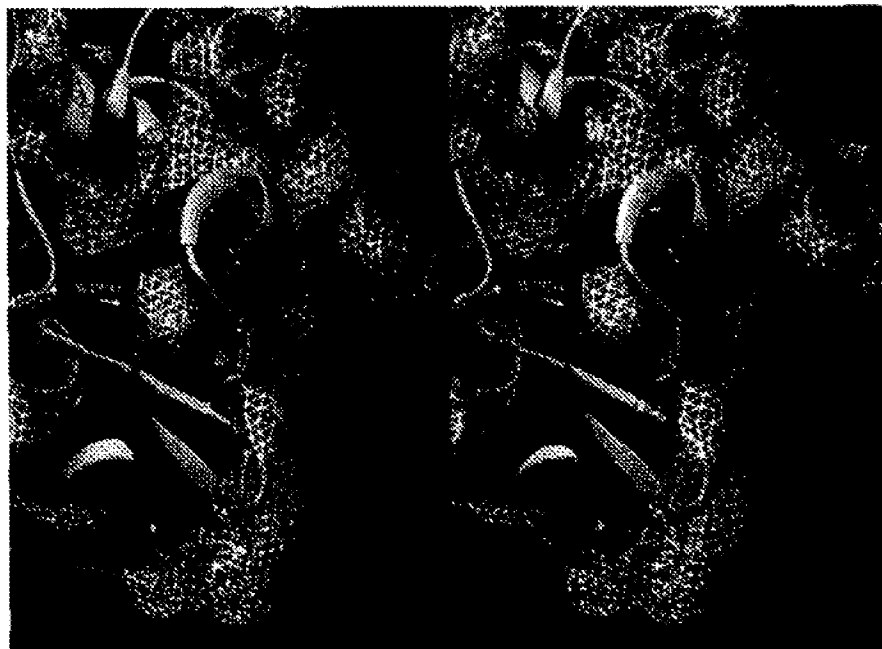
nificant interest, as such sites may display different compositions with respect to the degree of polarity, conformational flexibility and solvation characteristics. In addition, the lower affinity and specificity expected for ligand binding to artificial cavities should provide unique and challenging tests of computational predictions.

Detailed characterization of the interactions between ligand and protein is necessary before ligand binding to artificial cavities may be usefully compared to their natural counterparts. In previous studies of artificial protein cavities, two rather distinct types have been described. Largely hydrophobic buried cavities created in phage T4 lysozyme<sup>1,2,14,15</sup> generally do not appear to be specifically solvated, and display a mixture of rigid and conformationally mobile characteristics. Such cavities appear to bind hydrophobic ligands with a specificity that is determined by a combination of packing, desolvation, and conformational entropy factors. On the other hand, several polar buried cavities have been created in the enzyme cytochrome c peroxidase (CCP), each of which specifically bind polar ligands.<sup>3,4,16-22</sup> One such cavity, W191G, created by deletion of the side-chain of a tryptophan radical cation center is shown in Figure 1. In contrast to the hydrophobic cavities of T4 lysozyme, several molecules of solvent and a monovalent cation occupy the W191G cavity mutant of CCP in the absence of ligand, and the cavity was observed to bind small cationic heterocyclic compounds.<sup>3,4</sup> In many respects, the interactions seen in the polar W191G CCP cavity are more reminiscent of natural substrate-binding sites, with the potential for specific hydrogen bonding and electrostatic interactions. It has not been clear, however, how promiscuous the W191G cavity is with respect to ligand variation, whether more than one mode of binding is utilized, what range of affinities can be achieved, or how much the surrounding protein structure responds to ligand variation. Here, we report the structural and thermodynamic characterization of a variety of ligands bound to this cavity to address these questions.

## Results

### Specificity and thermodynamics of ligand binding

Initially, a wide variety of small organic molecules were screened for their ability to bind to W191G. Previous studies have shown that displacement of the solvent within the W191G cavity by ligands such as imidazole results in a subtle but



**Figure 1.** A stereo view of the W191G cavity mutant of CCP showing the protein backbone ribbon (brown) and the solvent-accessible surface (blue) calculated using a 1.4 Å probe sphere. The heme cofactor is shown in red, and the front surface of the molecule has been clipped off in order to show the interior of the protein. The buried W191G cavity is seen in the interior beneath the heme as labeled.

characteristic perturbation of the heme Soret absorption band.<sup>3</sup> Thus, samples of W191G were screened by titration against potential ligands selected primarily by their ability to fit into the known volume of the W191G cavity. As controls, identical titrations were carried out against WT CCP and the W191F mutant, the latter of which has heme Soret absorption features similar to those of W191G and can be considered to be a model for the cavity that is irreversibly occupied by the Phe side-chain.<sup>18,23</sup> These features of the W191F control were particularly useful in discriminating binding from non-binding ligands. Those compounds that produced the characteristic optical perturbation in W191G but not W191F were designated as potentially binding. As a result of this analysis it became apparent that previous conclusions<sup>3,4</sup> about the determinants for ligands to W191G were accurate; namely, that the cavity is specific for small heterocyclic cations. A number of these, such as imidazoles, exist in equilibrium between neutral and protonated cationic forms with physiologically accessible  $pK_a$  values. In each such case tested, ligands appeared to bind more tightly to W191G below their respective  $pK_a$  value than above it (data not shown). Based on these results, a second round of candidates were screened by focusing on shape-selected compounds with physiologically accessible cationic forms to produce a list of approximately 20 compounds that appear to bind specifically to W191G. From these screens, the following trends were readily apparent: (i) all tested

heterocyclic candidates that are small enough to fit into the cavity and are cationic under the experimental conditions elicited the optical binding response; (ii) neutral compounds, such as indole, or compounds with a  $pK_a$  well below the pH of the titration (typically pH 4.5), such as cyanopyridine or tetrazole, did not bind; and (iii) cationic compounds that are non-planar or that would require protein conformational change to alter the cavity dimensions, such as trimethylamine and pyrroline, gave no evidence in support of binding.

Dissociation constants were measured for compounds identified to bind W191G, using two independent methods. Measurement of the perturbation in the heme Soret absorption was used to obtain dissociation constants in titrations of W191G with ligands, and for selected compounds, binding parameters were obtained by isothermal titration calorimetry (ITC). These data are presented in Table 1, where entries that include enthalpies and entropies were derived from ITC measurements and the remaining entries were from optical titrations. In agreement with the initial screens and with previously published data,<sup>3,4,18,19</sup> it was clear that only compounds that were predominantly cationic at the measurement pH of 4.5 bound to the protein, while those that were neutral under the experimental conditions (i.e.  $pK_a$  below 4.5), were observed not to bind. Thus, all data presented in Table 1 have been corrected for the  $pK_a$  of the compound by taking into account the concentration in solution of the cationic form, so that

Table 1. Thermodynamic parameters of ligand-binding to W191G

ID	ligand	pK <sub>a</sub>	K <sub>d</sub> (mM)	ΔG (kcal/mol)	ΔH (kcal/mol)	ΔS (cal/molK)
2a5mt	2-Amino-5-methylthiazole	5.9	0.006 0.008	-7.1 -6.9	-14	-24
12dmi	1,2-Dimethylimidazole	7.9	0.03	-6.2		
2at	2-Aminothiazole	5.3	0.04 0.04	-6.0 -6.0	-13	-24
4ap	4-Aminopyridine	9.2	0.04 0.05	-6.0 -5.9	-23	-55
3ap	3-Aminopyridine	6.0	0.04 0.07	-6.0 -5.7	-17	-35
1mim	1-Methylimidazole	7.3	0.06 0.05	-5.8 -5.9	-4	4
2ap	2-Aminopyridine	6.9	0.05 0.07	-5.9 -5.7	-16	-38
34dmt	3,4-Dimethylthiazole	- <sup>a</sup>	0.05 0.11	-5.9 -5.4	-10	-15
4mim	4-Methylimidazole	7.5	0.06	-5.8		
im	Imidazole	7.0	0.07	-5.7		
2mim	2-Methylimidazole	7.6	0.07	-5.7		
imp	Imidazo[1,2-a]pyridine	>4.5	0.09	-5.5		
an	Aniline	4.6	0.03	-6.2		
py	Pyridine	5.2	0.14	-5.3		
bzi	Benzimidazole	5.5	0.15	-5.2		
345tmt	3,4,5-Trimethylthiazole	- <sup>a</sup>	0.20 0.13	-5.1 -5.3	-15	-34
ind	Indoline	4.7	0.16	-5.2		
1vim	1-Vinylimidazole	>5	0.15 0.14	-5.2 -5.3	-15	-30
2a4mt	2-Amino-4-methylthiazole	5.8	0.23 0.25	-5.0 -4.9	-12	-25
3mt	3-Methylthiazole	- <sup>a</sup>	0.30 0.11	-4.8 -5.4	-16	-37
quin	Quinoline	5.2	0.42	-4.6		
2eim	2-Ethylimidazole	8.0	0.73 0.78	-4.3 -4.2	-15	-36
234tmt	2,3,4-Trimethylthiazole	- <sup>a</sup>	1.5 0.56	-3.9 -4.4	-14	-33
tet	Tetrazole	0.6	nb <sup>b</sup>			
pz	Pyrazole	2.5	nb <sup>b</sup>			
3apz	3-Aminopyrazole	1.7	nb <sup>b</sup>			
2mp	2-Mercaptopyrimidine	1.2	nb <sup>b</sup>			
3cp	3-Cyanopyridine	3.6	nb <sup>b</sup>			
iso	Isoniazide	<4.5	nb <sup>b</sup>			
anth	Anthranilonitrile	2.1	nb <sup>b</sup>			
5a4pc	5-Amino-4-pyrazole carbonitrile	unk <sup>c</sup>	nb <sup>b</sup>			
4a5ic	carboxamide	unk <sup>c</sup>	nb <sup>b</sup>			
indz	Indazole	1.3	nb <sup>b</sup>			

Entries including values of ΔH and ΔS were obtained by isothermal titration calorimetry, while remaining entries were obtained from fits of binding titrations observed by the perturbation of the heme Soret absorbance. All values are corrected for the pK<sub>a</sub> of the ligand and represent the K<sub>d</sub> for the cationic forms. All measurements were performed at 25 °C in 100 mM Bis-Tris propane (pH 4.5) to favor protonation of the ligands. The buffer was adjusted without the use of inorganic salts to prevent competition from cations.<sup>4</sup>

<sup>a</sup> Ligand is cationic without protonation.

<sup>b</sup> No binding detected.

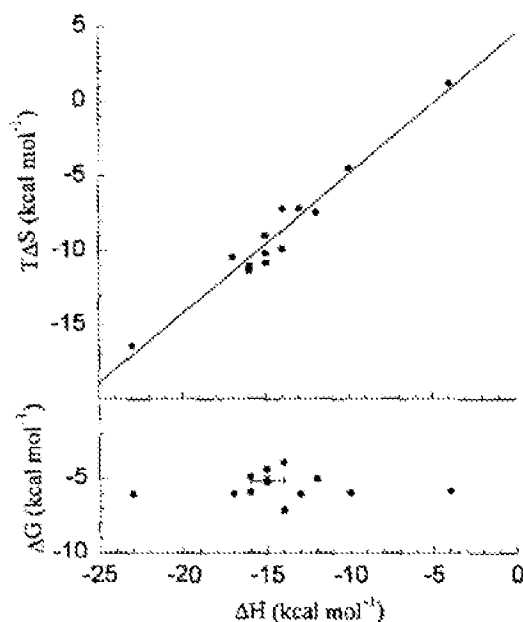
<sup>c</sup> Unknown.

these values represent intrinsic binding affinities of the cationic forms. Overall, the agreement between the optical and calorimetric data is good. Dissociation constants range from 6 μM for 2-amino-5-methylthiazole (2a5mt) to approximately 1 mM in the case of 2,3,4-trimethylthiazole (234tmt) for those ligands shown to bind. This represents a range of 3 kcal/mol (1 cal = 4.184 J) in binding free energy. Both the binding enthalpies and entropies were observed to be negative, as is usually observed in ligand-protein binding.<sup>29</sup> One significant observation is that there is a much greater

variation in binding enthalpy than for the total free energy, which implies a significant entropy-enthalpy compensation, as shown in Figure 2.

### Structural characterization of ligand binding

Ligand binding to the W191G cavity was characterized by X-ray diffraction using crystals soaked in mother liquor containing various ligands. Processed data sets (see Table 2) for ligand-soaked crystals were matched with previously determined structures of unsoaked W191G crystals of the



**Figure 2.** A plot of the thermodynamic parameters determined for ligand binding to the W191G cavity. Estimated error bars shown in the bottom graph for one of the ligands, 345tmt, were obtained from four repeated measurements. This plot illustrates that  $\Delta H$  values vary by more than the total free energy,  $\Delta G$ . Thus, a large entropy-enthalpy compensation is observed as shown in the above graph.

appropriate crystal form.<sup>3</sup> Ligand binding typically displaces the four well-ordered water molecules and potassium ion observed in the cavity in the absence of ligand, as shown in Figure 3(a). Thus, direct  $F_{\text{soak}} - F_{\text{unsoak}}$  difference maps were often not easily interpretable because both positive and negative density features are superimposed. Instead, a series of  $F_{\text{soak}} - F_c$  omit maps were created, where  $F_{\text{soak}}$  are the observed structure factors of the ligand-soaked crystal and  $F_c$  are the structure factors derived from a model for unsoaked W191G that was refined without any atoms or water molecules included in the cavity. Such omit maps represent the electron density of the contents of the cavity under the soaking conditions. These omit maps, contoured at  $4\sigma$  and  $8\sigma$ , are shown in Figure 3(b)-(s) for 18 of the ligands that were designated as potential cavity binders. Clearly interpretable density is observed in each case, indicating occupation by the specified ligand. Significantly, no difference density above about  $2\sigma$  was observed with any ligand in any other region of the structure remote from the W191G cavity, clearly demonstrating that all binding interactions are specific for the artificial cavity. This is in spite of the fact that several of the ligands might be capable of coordinating the heme iron by occupying the naturally occurring channel that is distal to the

heme (partially visible in Figure 1). In addition, test soaks with ligands that were designated as non-binders in the initial screens, for example tetrazole and 3-aminopyrazole, showed only the pattern associated with solvent in the unsoaked cavity (Figure 3(a)).

In almost all cases, it was possible to assign the orientation of ligands within the cavity. For three ligands, including 1,2-dimethylimidazole (12dmi), aniline, and 2-ethylimidazole (2eim), the shape of the electron density envelope was sufficient to assign ligand orientation. For others containing sulfur atoms in the ring, the position of the sulfur atom was evident in the electron density at a higher contour level, as shown in red in Figure 3. From this additional information, it was possible to assign ligand orientation for 2-aminothiazole (2at), 2-amino-4-methylthiazole (2a4mt), 3-methylthiazole (3mt), 3,4,5-trimethylthiazole (345tmt), and 2,3,4-trimethylthiazole (234tmt). Finally by combining the above information with inferences derived from hydrogen bonding interactions, as discussed below, orientations of 2-amino-5-methylthiazole (2a5mt), 3-aminopyridine (3ap), 1-methylimidazole (1mim), 2-aminopyridine (2ap), 3,4-dimethylthiazole (34dmt), 2-methylimidazole (2mim), imidazo[1,2-a]pyridine (imp), indoline, and 1-vinylimidazole (1vim) can be assigned unambiguously. Of the 18 ligands, 4-aminopyridine (4ap) cannot be assigned to a specific orientation, and 1vim contains insufficiently resolved density to assign the conformation of the vinyl group. These observations may result from lack of resolution, structural landmarks, or true conformational heterogeneity within the cavity. Nevertheless, reasonable estimates of the most likely orientation for these compounds were possible. Based on these assignments, models for the ligands were manually placed into the electron density as shown in Figure 3(b)-(s).

With few exceptions, ligand binding to W191G displaces the solvent of the ligand-free cavity without changes in the cavity wall. Previous studies have shown that W191G in the absence of ligand contains four water molecules, W401-W404, and one monovalent cation, either  $K^+$  or  $Na^+$  that are well ordered in the cavity and form a network of hydrogen bonds between themselves and with polar groups on the cavity walls.<sup>3,4</sup> These interactions are listed in the first row of Table 3. One water molecule, W308 shown in Figure 3a, was not introduced by the cavity but is observed in wild-type (WT) CCP and is retained in W191G. With the exception of 2at, 3ap, and aniline, each of the ligands displaces all of the solvent atoms introduced by cavity formation, without displacing the conserved W308. For 2at, 3ap, and aniline, one of the cavity water molecules (W401) remains in the cavity in addition to the ligand. While most ligands bind to W191G without inducing changes in the dimensions or shape of the cavity walls, two of the ligands, indoline and imp, appear to have induced an alternate protein conformation in the

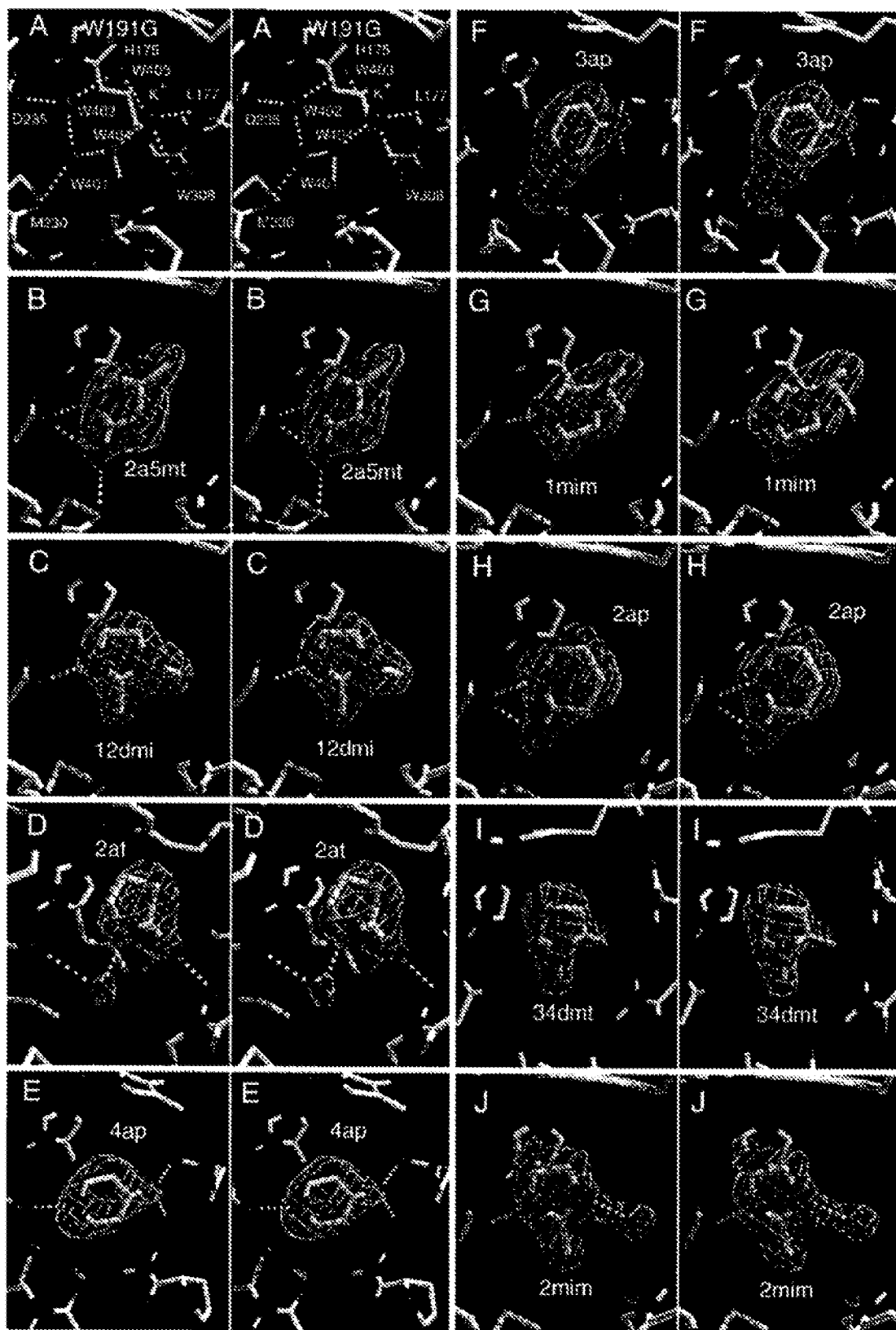


Figure 3 (legend shown on page 852)

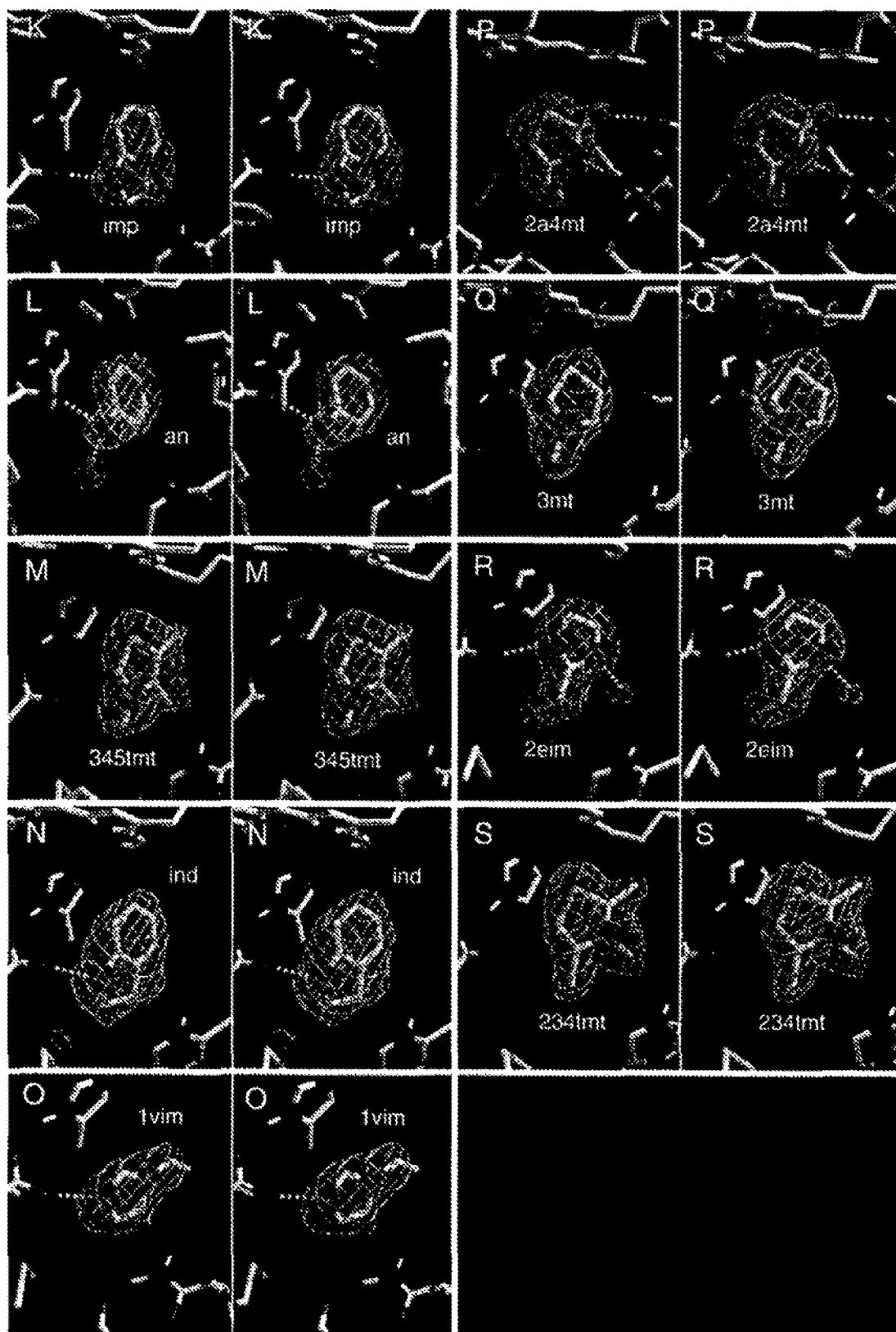


Figure 3 (legend shown on page 852)



cavity. For these two structures, significantly reduced electron density is seen in the protein for residues 190-195, indicating that this region of the structure has become mobile or is occupying multiple conformations.

Several modes of hydrogen bonding interactions are observed in the crystal structures of Figure 3. Hydrogen bonds inferred from the assigned ligand orientations are listed in Table 3, where the rows are sorted in order of the binding affinity, and the columns ordered by the number of total interactions. The most common hydrogen bond is to the carboxylate side-chain of Asp235. This interaction occurs for 12 of the 18 ligands and for two of them, 2a5mt and 2ap, Asp235 appears to make a bifurcated hydrogen bond to two atoms in the ligand. The second most often observed interaction between ligand and cavity, which occurs in six of the ligands, is a hydrogen bond to the conserved water molecule of WT CCP, W308. In addition, a few of the ligands make hydrogen bonds to the backbone carbonyl atom of L177, H175, and M230. Each of the three ligands, 2at, 3ap, and aniline, which retain cavity water W401 also form hydrogen bonds with it. Four of the ligands, 34dmt, 345mt, 3mt, and 234tmt, are methylthiazoles that are cations without traditional hydrogen bonding capability. However, these four compounds orient their most acidic CH proton (that adjacent to the endocyclic ring nitrogen) so that it is directed towards the Asp235 carboxylate group. Finally, it is noted that every atom in each of the 18 ligands that is capable of forming a hydrogen bond does so with some atom in the protein or included solvent.

## Discussion

The foremost conclusions from this study concern the unique specificity for ligand binding to the artificial cavity of W191G. Several cavities have been introduced into proteins by deletion of amino acid side-chains, and it is often observed that these cavities bind small molecules that mimic the properties of the removed structure.<sup>1-4,14-22</sup> Thus, artificial cavities can be utilized as templates for binding behavior. For example, several cases have been reported in which amines rescue function of lysine residues that have been replaced by smaller side-chains.<sup>23</sup> Other artificial cavities that have been created by removal of side-chains in the

interior of a protein have generally been non-polar in nature, reflecting the properties of the buried residues removed to create them.<sup>1,2,14,15</sup> These non-polar cavities do not appear to be specifically solvated, and bind hydrophobic ligands without making specific interactions. Thus, consideration of hydrogen bonding and electrostatic effects are minimized and the crucial concerns are with hydrophobic interactions, packing and conformational entropy. However, the cavity introduced by the W191G mutation of CCP is quite different in these respects. The Trp191 side-chain, while itself quite hydrophobic, is a special site in the enzyme that is oxidized reversibly to a stable radical cation form as a part of the enzymatic mechanism.<sup>26</sup> Not surprisingly, the structure surrounding this site has been evolved specifically to stabilize the cation form.<sup>4,27-30</sup> It is remarkable that the ligand-binding specificity of the W191G cavity faithfully reflects this evolved property, and that all of the 20 compounds identified to date that bind to W191G are cationic heterocycles that bind only to the cavity by making an array of hydrogen bonding interactions.

Subsets of the same interactions that are believed to stabilize the naturally occurring Trp191 cation of WT CCP are utilized to define the specificity determinant of W191G. We conclude that electrostatic interactions with Asp235 and several backbone carbonyl atoms define the cation requirement,<sup>4,31,32</sup> while specific hydrogen bonding interactions to these groups determine the ligand orientation within the cavity. An important observation is that the highly ordered solvent observed in the W191G cavity in the absence of ligand forms an ideal template for ligands, because all potential hydrogen bonds to the cavity walls are satisfied. While none of the discovered ligands makes all of these potential hydrogen bonds simultaneously, no new hydrogen bonds are utilized by ligands that were not also seen with solvent alone. Of the six different hydrogen bonds made to ligands, Asp235 clearly dominates and appears to determine ligand orientation. This is seen from comparisons of the isosteric ligands 2ap, 3ap, 4ap, and aniline, where each ligand adopts a different orientation in order to permit a hydrogen bond with Asp235. Similarly, 2eim and 1vim orient their substituents differently within the cavity in order to preserve interaction with Asp235. Finally, 34dmt, 345mt, 3mt, and 234tmt, are cationic methylthiazoles without tra-

**Figure 3.** A series of stereo electron density omit maps showing the contents of the W191G cavity after soaking crystals in a variety of ligands. Shown in (a) is the solvent occupation of the cavity for a crystal without soaking in ligand. Positions of four water molecules, W401-W404, and a potassium ion are indicated that fill the cavity created by removal of the Trp191 side-chain. Also labeled are protein residues that are involved in hydrogen bonding interactions with the occupied solvent and a water molecule, W308, which is observed in WT CCP and is retained in the W191G structure. Shown in (b)-(s) are omit electron density maps contoured at 4 $\sigma$  (white) and 8 $\sigma$  (red) after soaking crystals in the indicated ligand (see Table 1 for description of the ligands). Models of the ligand are also shown placed in the observed electron density. The ligands are presented in the order of the relative binding affinity as shown in Table 1.

Table 2. Crystallographic data collection statistics

Ligand	2aim	lvim	an	2ap	3ap	4ap	3ut	34dnt	2a5mt	2a4mt	ind	imp
PDB code	laeq	laej	laee	laeo	laef	laeg	laeb	laed	laen	laeh	laek	laem
Crystal form*	1	1	1	2	1	1	1	1	2	1	1	1
Cell <i>a</i> , <i>b</i> , <i>c</i> (Å)	107.4	107.6	108.0	103.6	106.6	106.3	106.1	107.3	105.3	107.3	108.8	106.6
	76.6	77.1	77.2	73.3	75.9	75.7	75.8	76.5	74.2	76.5	76.8	76.0
	51.5	51.3	54.1	44.4	51.1	51.1	51.1	51.3	45.5	51.6	52.0	51.3
Resolution (Å)	2.0	2.1	2.0	2.2	2.0	2.0	2.1	2.2	2.1	1.9	2.1	2.0
<i>I</i> / <i>σ<sub>i</sub></i> (av)	12.9	8.4	15.7	10.2	11.9	14.9	15.4	10.2	13.6	24.2	14.7	14.5
<i>I</i> / <i>σ<sub>i</sub></i> (last shell)	1.65	0.49	2.67	1.73	1.09	1.39	1.71	1.80	1.12	1.91	1.45	1.46
No. reflections	30,561	22,158	26,238	20,946	23,265	27,229	22,213	20,736	16,307	29,343	17,064	27,849
Complete (%)	95	84	96	90	79	89	89	76	73	87	64	92
<i>R</i> <sub>int</sub>	0.051	0.085	0.047	0.107	0.087	0.076	0.067	0.092	0.073	0.069	0.060	0.065

Each of the data sets was obtained at 18 °C from a single crystal soaked in the indicated ligand as described in Experimental Methods. Structures of W191G in the absence of ligand (laad) and with 1mm, 2mm, 12mm (lump), 2at (laev), 234mt (laed) and 345mt (laef)<sup>a</sup> bound have been published.

\* Crystals were observed to occur in one of two forms that differ in packing orientation within the asymmetric unit as previously described.<sup>2</sup> Form 1 was previously referred to as the yeast form and form 2 as the mkt form based upon the protein source where the form was first observed.

Table 3. Hydrogen bonding interactions between the W191G cavity and bound ligands

Ligand	D235 (OD2)	W308	L177 (O)	H175 (O)	W401	M230 (O)	G191 (NH)	G189 (O)
K400, W401-W404	X	X	X	X		X	X	X
2a5mt	X					X		
12dmj	X							
2at		X		X	X			
4ap	X	X	X					
3ap		X	X		X			
1mim	X							
2ap	X							
34dmt								
2mim	X	X						
imp	X							
an	X				X			
345tmt								
ind	X							
1vim	X							
2a4mt			X	X				
3mt								
2eim	X	X						
234tmt								

Hydrogen bonding interactions are indicated for the assigned conformations shown in Figure 3 having hetero-atom distances between 2.5 and 3.4 Å. Ligands are shown in order of decreasing affinity.

ditional hydrogen bonding capability. However, each of these four compounds orients its most acidic CH proton (that adjacent to the endocyclic ring nitrogen) so that it is directed towards the Asp235 carboxylate group. This interaction, discussed previously in terms of an unconventional CH to O hydrogen bond,<sup>19</sup> illustrates the remarkable ability of the ligand-cavity interactions to mimic hydrogen bonding patterns specified by the template of the removed tryptophan side-chain.

Calorimetry data show that ligand binding to W191G is enthalpy-driven, but the overall binding free energy is relatively weak, does not vary greatly, and exhibits large enthalpy-entropy compensation. These factors make empirical correlations of the relative binding energies difficult to predict without detailed calculations. Entropy-enthalpy compensation is commonly observed in aqueous solution, and is usually characterized by slopes near unity,<sup>33</sup> as observed in this case. This effect is often attributed to a trade-off in which a stronger enthalpic interaction results in a compensating loss of entropy due to motional restriction.<sup>34</sup> It is expected that the solvation energies of the ligands used in this study will be quite similar. Indeed, calculations using MEAD<sup>35</sup> (data not shown) indicate a variation of only about 20% in the solvation energies of these compounds. Similarly, no readily apparent relationship exists between the number or type of hydrogen bonds made to the ligand and the relative thermodynamic parameters. This is not unexpected, as these very polar compounds will be fully hydrogen bonded in solution, and have also found fully satisfied hydrogen bond partners within the cavity. Thus, while hydrogen bonding interactions are expected to contribute large absolute interaction enthalpies with the cavity and may dictate ligand orientation, differences between bound and

unbound interactions will be similar for each compound. A more sophisticated attempt was made to correlate electrostatic interactions with binding enthalpies. As described previously, electrostatic calculations using POLARIS and the PDLD (protein dipoles Langevin dipoles) method<sup>32,36</sup> were performed to estimate the electrostatic interactions between the ligand and the surrounding protein. These calculations (not shown) also failed to account for variations in the observed enthalpies or free energies. The small difference (~3 kcal/mol) in binding free energy between the tightest and weakest of these ligands most likely reflects a combination of subtle interactions including electrostatics, hydrogen bonding, desolvation entropies, and packing interactions that must each be accounted for very accurately in order to predict relative affinities. Previously successful applications of the PDLD method in comparing the relative binding of 234tmt and 345tmt<sup>19</sup> or in the comparison of the relative stabilization of the tryptophan cation radical in CCP relative to ascorbate peroxidase<sup>32</sup> likely resulted from the cancellation of elements such as solvation and non-polar interactions that are not well modeled by the specific implementation of the PDLD method used in these studies.

As with the previously studied cavities in T4 lysozyme mutants, the W191G cavity contains regions that appear structurally rigid, and others that are subject to movement. It might be expected that artificially created cavities will undergo some collapse, or be more conformationally mobile than natural binding sites. However, a detailed analysis of cavity collapse has been reported for T4 lysozyme mutants,<sup>1,15</sup> and it is remarkable that the cavities that have been characterized to date generally appear quite rigid. One such cavity, the L99A cavity of T4 lysozyme, was observed to have a region

that is structurally rigid and another region that deforms in response to packing interactions with ligands.<sup>14</sup> In the case of the W191G cavity of CCP, very little, if any, collapse or movement in the dimensions of the cavity is seen upon binding for most of the ligands. However, the cavity is inaccessible to solvent in the average crystal structure, and thus the relatively rapid exchange with ligands requires a significant role for conformational dynamics in order to allow ligand access.<sup>17</sup> Indeed, when benzimidazole is bound to W191G, a large loop rearrangement between Pro190 and Asn195 has been observed to give an open channel conformation, providing a tantalizing view of the pathway for ligand access.<sup>17</sup> It is of interest that the two largest ligands of this study, imp and indoline, show significantly reduced electron density for the protein at this same region, suggesting conformational heterogeneity of this loop. Thus, this cavity appears to contain a specific hinged gate with open and closed conformations to allow ligand access, while the remaining walls of the cavity appear quite rigid and resistant to either collapse or induced fit binding behavior.

The specific ligand-binding properties of the W191G cavity may help establish practical uses for engineered cavities. For example, a number of factors make the W191G cavity an ideal test bed for use in the evaluation and development of computational tools for drug design. The rigidity of the cavity wall upon binding ligands simplifies the task of accounting for cavity collapse or adjustment in the protein structure. The polar nature of the cavity and its ligands provide a template that is similar to naturally occurring enzyme sites. The small variation in binding free energies provides a challenging test for calculations of the delicately balanced forces involved. Finally, the crystal structures of a complete series of ligands bound to the same site along with the thermodynamic parameters provides a well-defined framework for such calculations. Computational studies of ligand-binding energetics have already made use of the W191G cavity mutant for method development. These studies have resulted in improved computational methodologies and have provided specific predictions concerning potential conformational heterogeneities for certain ligands that are testable by further experiments.<sup>37,38</sup>

## Experimental methods

### Protein expression and purification

W191G apoprotein was prepared, reconstituted with heme, and purified as described.<sup>3</sup> Purified enzyme was recrystallized twice from distilled water and stored as a crystal suspension at 77 K. Before use, crystals were washed in cold distilled water and dissolved in the appropriate buffer. W191G protein concentrations were determined using  $\epsilon_{412} = 104 \text{ mM}^{-1} \text{ cm}^{-1}$  as determined from the pyridine hemochromogen assay.

### Ligand-binding measurements

Two methods were used to characterize ligand binding to the W191G cavity. Simple dissociation constants were obtained by optically detected titrations as described.<sup>3</sup> These measurements are based upon a small perturbation of the heme Soret absorbance that is observed when ligands displace the solvent in the W191G cavity. For these measurements, ligand stock solutions were prepared in 50% (v/v) ethanol, except for indoline, imidazo[1,2-a]pyridine, and quinoline which were prepared in 95% ethanol. Prior to use, each stock solution was adjusted to the same pH as the protein buffer with  $\text{H}_3\text{PO}_4$ . Dissociation constants were determined from Scatchard plots based on the difference absorbance of the Soret maximum, assuming one binding site per protein molecule. For selected ligands, a more extensive characterization of the ligand-binding thermodynamics was obtained by isothermal titration calorimetry (ITC). These measurements were made using an MC2 titration calorimeter from Microcal, Inc. at 25°C in 100 mM Bis-Tris propane at pH 4.5. Protein solutions of known concentration were titrated with known concentrations of ligand (typically 30-fold higher concentration relative to protein) by automated 5  $\mu\text{l}$  injections from a 100  $\mu\text{l}$  syringe (400 rpm) at intervals of four minutes. Titrations were extended past the end-point to allow subtraction of the heat of ligand dilution from the injection peaks. Data analysis was performed using ORIGIN software customized for ITC analysis by fitting the data to a single-site binding isotherm to obtain values for the ligand association constant, molar enthalpy, and the number of ligand sites. Binding free energies were calculated as  $-RT \ln K_d$ , using a 1 M standard state, and binding entropies were calculated as  $(\Delta G - \Delta H)/T$ . With the exception of 3,4-dimethylthiazole and 2,3,4- and 3,4,5-trimethylthiazoles, which were synthesized by known methods,<sup>39</sup> all ligands were obtained from Aldrich. As the results show that compounds bind only in a cationic form (see Results), each of the reported values for binding constants and free-energies has been corrected where appropriate for the  $\text{pK}_a$  of the ligand to reflect the intrinsic affinity of the cationic form.

### X-ray crystallographic data collection and analysis

Crystals of W191G were grown for X-ray data collection by vapor diffusion at 18°C from sitting drops of CCP in 8.5% (v/v) 2-methyl-2,4-pentanediol (MPD), 200 mM Mes (pH 6.0), against a reservoir of 25% MPD, as described.<sup>16</sup> Crystals were soaked in artificial mother liquor containing 30–50 mM ligand for 60 minutes before mounting and data collection. Diffraction data were collected at 15°C using CuK $\alpha$  radiation from the rotating anode of a Siemens SRA X-ray generator using a Siemens area detector. Data were indexed and integrated using the XENGEN programs.<sup>40</sup> All data were analyzed by difference Fourier techniques using the Scripps Xtal-View software.<sup>41</sup> Models for the ligands were obtained by geometry optimization using density functional methods implemented in Gaussian-94, as described.<sup>19</sup>

### Protein Data Bank accession codes

The crystallographic coordinates for the structures presented in this work have been deposited with the RCSB Protein Data Bank (<http://www.rcsb.org>) with accession codes 1aeq, 1aej, 1aee, 1aef, 1aeg, 1aeb, 1aed, 1aen, 1aeh, 1aek, and 1aem.

## Acknowledgments

The authors thank Dr Duncan McRee, Professor C. L. Brooks III, Professor A. A. Olson, Dr David Goodsell and Dr Garret Morris for helpful discussions. This work was supported in part by grant GM41049 from the National Institutes of Health to D.B.G., predoctoral fellowships from the DOE and the La Jolla Interfaces in Sciences to R.J.R., and by an NSRA NIH fellowship GM17844 to R.A.M.

## References

- Eriksson, A. E., Baase, W. A., Zhang, X. J., Heinz, D. W., Elabier, M., Baldwin, E. P. & Matthews, B. W. (1992). Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science*, **255**, 178-183.
- Eriksson, A. E., Baase, W. A., Wozniak, J. A. & Matthews, B. W. (1992). A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene. *Nature*, **355**, 371-373.
- Fitzgerald, M. M., Churchill, M. J., McRee, D. E. & Goodin, D. B. (1994). Small molecule binding to an artificially created cavity at the active site of cytochrome *c* peroxidase. *Biochemistry*, **33**, 3807-3818.
- Fitzgerald, M. M., Trester, M. L., Jensen, G. M., McRee, D. E. & Goodin, D. B. (1995). The role of aspartate-235 in the binding of cations to an artificial cavity at the radical site of cytochrome *c* peroxidase. *Protein Sci.* **4**, 1844-1850.
- Kuntz, I. D. (1992). Structure-based strategies for drug design and discovery. *Science*, **257**, 1078-1082.
- Clackson, T. & Wells, J. A. (1995). A hot spot of binding energy in a hormone-receptor interface. *Science*, **267**, 383-386.
- Petrassi, H. M., Klabunde, T., Sacchetti, J. & Kelly, J. W. (2000). Structure-based design of *N*-phenyl phenoxazine transilthyretin amyloid fibril inhibitors. *J. Am. Chem. Soc.* **122**, 2178-2192.
- Ringe, D. (1995). What makes a binding site a binding site? *Curr. Opin. Struct. Biol.* **5**, 825-829.
- Pearlman, D. A. (1994). A comparison of alternative approaches to free energy calculations. *J. Phys. Chem.* **98**, 1487-1493.
- Radmer, R. J. & Kollman, P. A. (1997). Free energy calculation methods: a theoretical and empirical comparison of numerical errors and a new method for qualitative estimates of free energy changes. *J. Comput. Chem.* **18**, 902-919.
- Kong, X. & Brooks, C. L., III (1996). Lambda-dynamics: a new approach to free energy calculations. *J. Chem. Phys.* **105**, 2414-2423.
- Kuntz, I. D., Meng, E. C. & Shoichet, B. K. (1994). Structure-based strategies for drug design and discovery. *Acc. Chem. Res.* **27**, 117-123.
- Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. R., Cheatham, T. E., III, DeBolt, S., Ferguson, D. et al. (1995). AMBER, a computer program for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to elucidate the structures and energies of molecules. *Comp. Phys. Commun.* **91**, 1-41.
- Morton, A. & Matthews, B. W. (1995). Specificity of ligand binding in a buried non-polar cavity of T4 lysozyme: linkage of dynamics and structural plasticity. *Biochemistry*, **34**, 8576-8588.
- Morton, A., Baase, W. A. & Matthews, B. W. (1995). Energetic origins of specificity of ligand binding in an interior non-polar cavity of T4 lysozyme. *Biochemistry*, **34**, 8564-8575.
- McRee, D. E., Jensen, G. M., Fitzgerald, M. M., Siegel, H. A. & Goodin, D. B. (1994). Construction of a bisquo heme enzyme and binding by exogenous ligands. *Proc. Natl Acad. Sci. USA*, **91**, 12847-12851.
- Fitzgerald, M. M., Musah, R. A., McRee, D. E. & Goodin, D. B. (1996). A ligand-gated, hinged loop rearrangement opens a channel to a buried artificial protein cavity. *Nature Struct. Biol.* **3**, 626-631.
- Musah, R. A. & Goodin, D. B. (1997). Introduction of novel substrate oxidation into a heme peroxidase by cavity complementation: oxidation of 2-aminothiazole and covalent modification of the enzyme. *Biochemistry*, **36**, 11665-11674.
- Musah, R. A., Jensen, G. M., Rosenfeld, R. J., Bunte, S. W., McRee, D. E. & Goodin, D. B. (1997). Variation in strength of a CH to O hydrogen bond in an artificial cavity. *J. Am. Chem. Soc.* **119**, 9083-9084.
- Hirst, J. & Goodin, D. B. (2000). Unusual oxidative chemistry of *N*-hydroxyarginine and *N*-hydroxyguanidine catalyzed at an engineered cavity in a heme peroxidase. *J. Biol. Chem.* **275**, 8582-8591.
- Hirst, J., Wilcox, S. K., Williams, P. A., Blankenship, J., McRee, D. E. & Goodin, D. B. (2001). Replacement of the axial histidine ligand with imidazole in cytochrome *c* peroxidase. 1. Effects on structure. *Biochemistry*, **40**, 1265-1273.
- Hirst, J., Wilcox, S. K., Ai, J., Moenne-Loccoz, P., Loefer, T. M. & Goodin, D. B. (2001). Replacement of the axial histidine ligand with imidazole in cytochrome *c* peroxidase. 2. Effects on heme coordination and function. *Biochemistry*, **40**, 1274-1283.
- Mauro, J. M., Fishel, L. A., Hazzard, J. T., Meyer, T. E., Tollin, G., Cusanovich, M. A. & Kraut, J. (1988). Tryptophan-191 > phenylalanine, a proximal-site mutation in yeast cytochrome *c* peroxidase that strongly affects the kinetics of ferrocyanochrome *c* oxidation. *Biochemistry*, **27**, 6243-6256.
- Koss, P. D. & Subramanian, S. (1981). Thermodynamics of protein association reactions: forces contributing to stability. *Biochemistry*, **20**, 3096-3102.
- Toney, M. D. & Kirsch, J. F. (1989). Direct Brønsted analysis of the restoration of activity to a mutant enzyme by exogenous amines. *Science*, **243**, 1485-1488.
- Sivaraja, M., Goodin, D. B., Smith, M. & Hoffman, B. M. (1989). Identification by ENDOR of Trp191 as the free-radical site in cytochrome *c* peroxidase compound ES. *Science*, **245**, 738-740.
- Goodin, D. B. & McRee, D. E. (1993). The Asp-His-Fe triad of cytochrome-*c* peroxidase controls the reduction potential, electronic structure, and coupling of the tryptophan free radical to the heme. *Biochemistry*, **32**, 3313-3324.
- Bonagura, C. A., Sundaramoorthy, M., Bhaskar, B. & Poulos, T. L. (1999). The effects of an engineered cation site on the structure, activity, and EPR properties of cytochrome *c* peroxidase. *Biochemistry*, **38**, 5538-5545.
- Pappa, H., Patterson, W. R. & Poulos, T. L. (1996). The homologous tryptophan critical for cytochrome *c* peroxidase function is not essential for ascorbate peroxidase activity. *J. Inorg. Biochem.* **1**, 61-66.

30. Bonagura, C. A., Sundaramoorthy, M., Pappa, H. S., Patterson, W. R. & Poulos, T. L. (1996). An engineered cation site in cytochrome *c* peroxidase alters the reactivity of the redox active tryptophan. *Biochemistry*, **35**, 6107-6115.
31. Miller, M. A., Han, G. W. & Kraut, J. (1994). A cation binding motif stabilizes the compound I radical of cytochrome *c* peroxidase. *Proc. Natl Acad. Sci. USA*, **91**, 11118-11122.
32. Jensen, G. M., Bunte, S. W., Warshel, A. & Goodin, D. B. (1998). Energetics of cation radical formation at the proximal active site tryptophan of cytochrome *c* peroxidase and ascorbate peroxidase. *J. Phys. Chem. ser. B*, **102**, 8221-8228.
33. Bundle, D. R. & Sigurskjold, B. W. (1994). Determination of accurate thermodynamics of binding by titration microcalorimetry. *Methods Enzymol.* **247**, 288-305.
34. Sigurskjold, B. W., Altman, E. & Bundle, D. R. (1991). Sensitive titration microcalorimetric study of the binding of Salmonella O-antigenic oligosaccharides by a monoclonal antibody. *Eur. J. Biochem.* **197**, 239-246.
35. Osapay, K., Young, W. S., Bashford, D., Brooks, C. L., III & Case, D. A. (1996). Dielectric continuum models for hydration effects on peptide conformational transitions. *J. Phys. Chem.* **100**, 2698-2705.
36. Jensen, G. M., Warshel, A. & Stephens, P. J. (1994). Calculation of the redox potentials of iron-sulfur proteins: the 2-/3- couple of [Fe4S4(Cys(4))] clusters in *Peptococcus aerogenes* ferredoxin, *Azotobacter vinelandii* ferredoxin I, and *Chromatium vinosum* high-potential iron protein. *Biochemistry*, **33**, 10911-10924.
37. Banba, S. & Brooks, C. L., III (2000). Free energy screening of small ligands binding to an artificial protein cavity. *J. Chem. Phys.* **113**, 3423-3433.
38. Banba, S., Guo, Z. & Brooks, C. L., III (2000). Efficient sampling of ligand orientations and conformations in free energy calculations using the lambda-dynamics method. *J. Phys. Chem.* **104**, 6903-6910.
39. Chen, Y. T. & Jordan, F. (1991). Reactivity of the thiazolium C2 ylide in aprotic solvents: novel experimental evidence for addition rather than insertion reactivity. *J. Org. Chem.* **56**, 5029-5038.
40. Howard, A. J., Nielson, C. & Xuong, N. H. (1985). Software for a diffractometer with multiwire area detector. *Methods Enzymol.* **114**, 452-472.
41. McRee, D. E. (1992). A visual protein crystallographic software system for X11/Xview. *J. Mol. Graph.* **10**, 44-46.

Edited by R. Huber

(Received 20 August 2001; received in revised form 16 November 2001; accepted 19 November 2001)